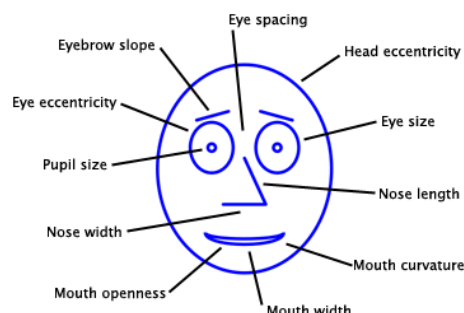




Chernoffovy tváře (ikonové grafy)

Jedním ze základních pomocníků pro datovou analýzu je vizualizace dat. Pokud chcete vizualizovat v jednom grafu jednu nebo dvě proměnné (jednu nebo dvě dimenze najednou), pak to není žádný problém, máte na výběr z mnoha druhů grafů a vizualizačních technik. V případě nutnosti podívat se na trojrozměrná data máme stále možnost situaci krásně analyzovat díky třírozměrným grafům s možností podívat se na rozložení bodů z více úhlů. Pokud ale pracujete ve svých datech s větším množstvím proměnných, začnete s klasickými grafy narážet. Schválně, jak bychom vykreslili čtyřrozměrný bodový graf, když jsme schopni vidět a myslet nanejvýš v našich třech rozměrech?

Nutnost vizualizovat takováto vícedimenzionální data vedla k vývoji jiných typů grafů. Jedním z nich zde začneme. Ukážeme si oblíbený typ vizualizace, kterou navrhl pan Herman Chernoff již v roce 1973. Jedná se o tzv. Chernoffovy tváře, v tomto grafu či vizualizaci je pro každý případ vykreslen jeden obličej, přičemž parametry obličeje jako například délka nosu, velikost uší, tvar hlavy, atd. jsou vykresleny podle hodnot jednotlivých proměnných. Nápad použít tváře pochází z myšlenky, že člověk je zvyklý a schopný jednoduše rozpoznat podobnosti nebo odlišnosti tváře. Ve *STATISTICE* je možné zaznamenat až 20 proměnných v jedné tváři.



Pomocí těchto grafů lze identifikovat případy se stejným chováním, rozdělovat do skupin, ale také nalézt případy s chováním výrazně odlišným od ostatních. To je možná ještě důležitější, ikonovým grafem dokážeme vizuálně odhalit odlehle případy (tj. případy s výrazně odlišnými hodnotami než má většina dat). Není nutno zmiňovat, že poznat hodnotu mimo většinu dat ve vícedimenzionálním prostoru není triviální – odlehlost v jedné z dimenzí je bezproblémová, nicméně to, že je hodnota ve všech rozměrech ve správných mezích ještě neznamená, že nebude odlehlá. Dalším speciálním využitím je zkoumání trendů a vůbec jakási možnost zažít si nějakým rozumnějším způsobem mnohorozměrná data.

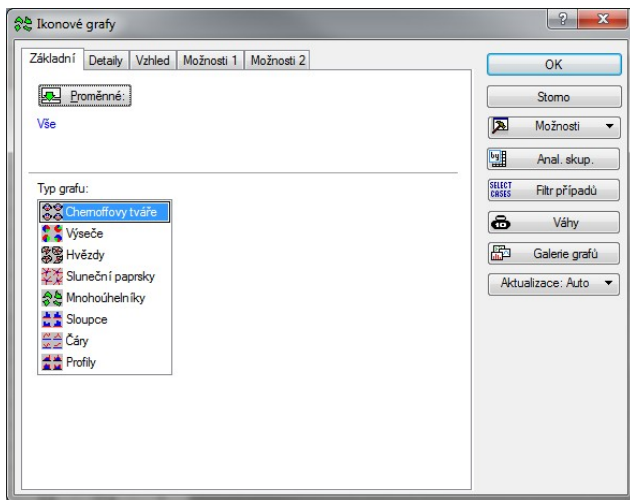
Obecně patří Chernoffovy tváře mezi takzvané ikonové grafy. V programu *STATISTICA* najdete ikonové grafy jednoduše přímo v záložce **Grafy**.

Dost teorie, teď bychom se vrhli na praktický příkládek.

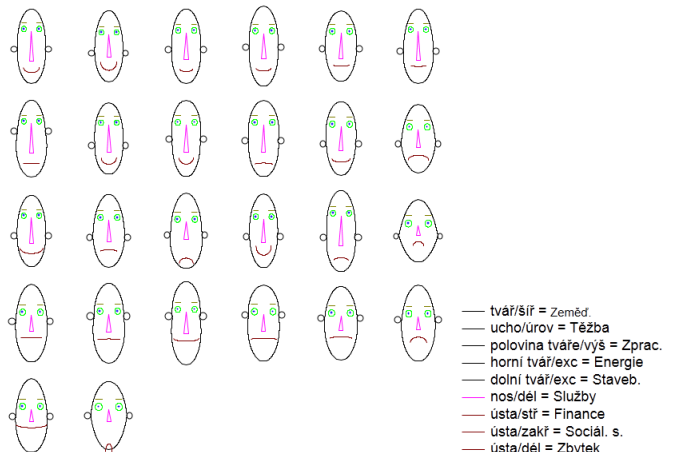
Příklad (Typ zaměstnání v jednotlivých zemích Evropy)

Pojďme na to. Vykreslíme si nejprve nějaké Chernoffovy obličeje na datech, která zachycují procentuální zastoupení zaměstnanců v jednotlivých sektorech v různých státech Evropy. Data můžete nalézt ke stažení [zde](#). Máme 9 proměnných, které chceme všechny najednou vizualizovat. Otevřeme si dialog ikonových grafů a zvolíme všech 9 proměnných.

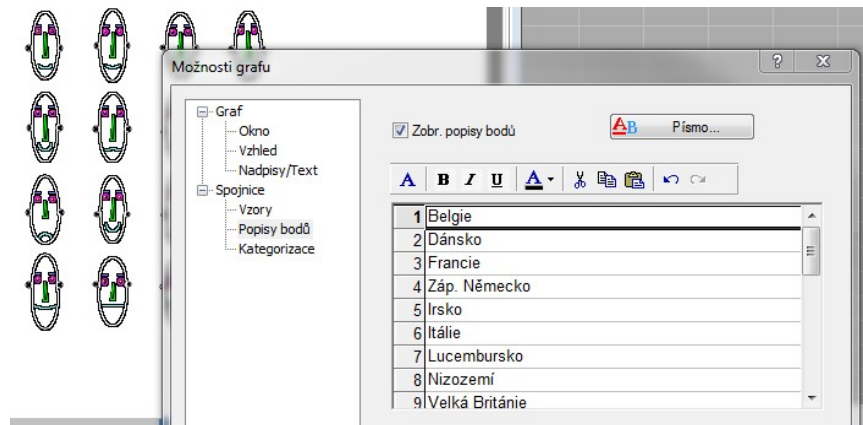
	Podíl pracovníků (%) zaměstnaných v jednotlivých oblastech hospodářství								
	1 Zeměd.	2 Těžba	3 Zprac.	4 Energie	5 Staveb.	6 Služby	7 Finance	8 Sociál. s.	9 Zbytek
Belgie	3,3	0,9	27,6	0,9	8,2	19,1	6,2	26,6	7,2
Dánsko	9,2	0,1	21,8	0,6	8,3	14,6	6,5	32,2	7,1
Francie	10,8	0,8	27,5	0,9	8,9	16,8	6	22,6	5,7
Záp. Německo	6,7	1,3	35,8	0,9	7,3	14,4	5	22,3	6,1
Irsko	23,2	1	20,7	1,3	7,5	16,8	2,8	20,8	6,1
Itálie	15,9	0,6	27,6	0,5	10	18,1	1,6	20,1	5,7

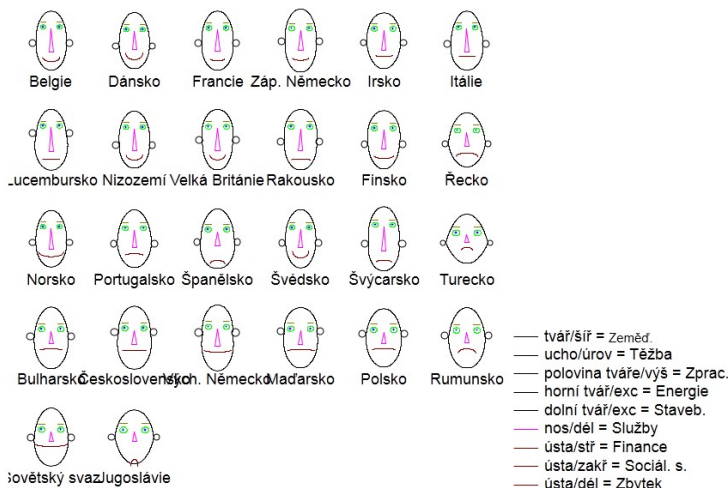


Klepeme **OK** a máme ihned výsledek:



Zdají se Vám některé „ksichtíky“ podobné? Nebo jsou zde naopak takové, které jsou naprosto odlišné od ostatních? Pokud jste si obličeje dostatečně prohlédli, jistě máte nějaké adepty. Určitě zde jsou takové, které se liší od ostatních například tím, že se prostě tváří mnohem divněji než ostatní. Tak to Vás teď bude jistě zajímat, který obličej přísluší kterému řádku (státu). Graf Chernoffových tváří je ve *STATISTICE* graf jako každý jiný, když tedy vyvoláte dialog **Možnosti grafu**, najdete možnost **Zobrazit popisy bodů**, kterou je potřeba zaškrtnout.



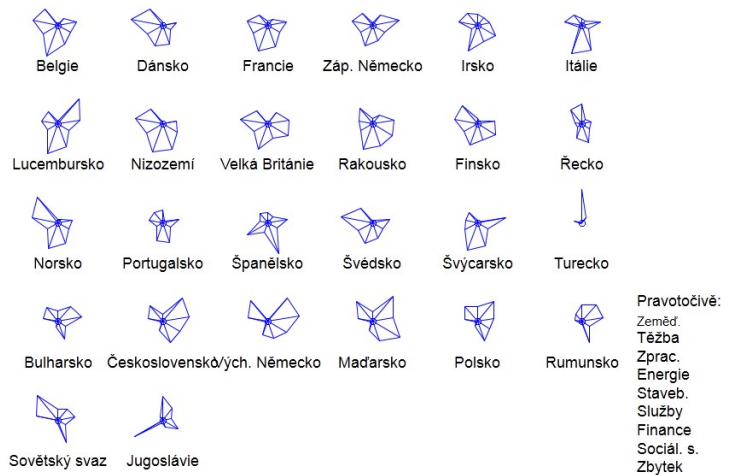


Nyní již vidíme, či je která tvář, pojďme se tedy podívat na výsledky. Tak schválně, které státy mají nejvíce odlišné obličej? Jasným kandidátem je Turecko a Jugoslávie, případně Řecko nebo Sovětský svaz. Téměř identicky vypadají například Belgie a Francie.

V legendě napravo od obličejů se poté můžete podívat, jak jsou jednotlivé proměnné do obličejů zakódovány.

Chernoffovy tváře jsou samozřejmě jen jednou formou ikonových grafů, zkuste si vyzkoušet také další typy jako třeba hvězdy.

Zde se každá dimenze (kerou je standardizovaná veličina) vykresluje postupně ve zvoleném pořadí dokola, délka příslušného cípu hvězdy je podle velikosti příslušné proměnné. Opět, možná dokonce více než předtím vidíme odlišnost Turecka a Jugoslávie. Pokud bychom hledali podobné hvězdy, pak lze jistě říct, že například Francie a Belgie mají podobné hvězdy, což znamená podobné rozvrstvení pracujících v jednotlivých sektorech. Taktéž například Norsko, Švédsko a Finsko mají velmi podobný charakter, což bychom čekali.

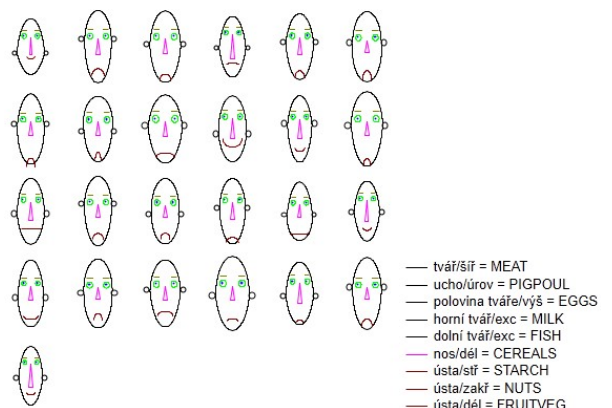


Domácí úloha

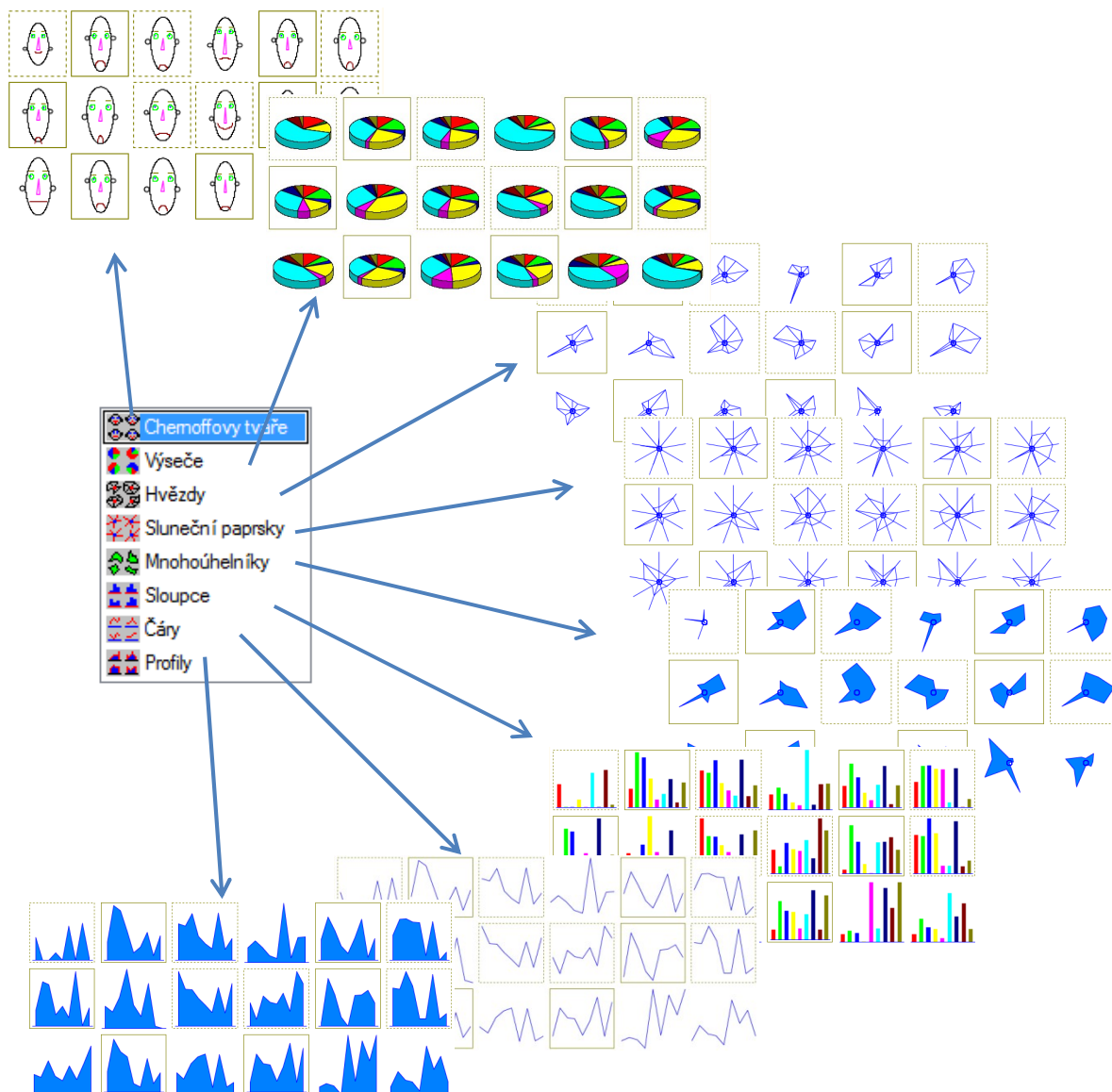
Pokud si chcete tento graf vyzkoušet sami na jiném datovém souboru, pak můžeme doporučit například soubor **Proteins.sta** (je k dispozici v příkladech programu **STATISTICA**, tedy **Soubor-Otevřít příklady-Datasets-Proteins.sta**). V tomto případě se jedná o odhady příslušnosti proteinů z 9 různých zdrojů pro obyvatele 25 zemí Evropy. Data pocházejí ze studie z roku 1973.

Vpravo je k dispozici náhled výsledků.

Pokud jste zvědaví a chcete zjistit, které tváře patří ke kterým státům, nezbyde Vám nic jiného než si tento příklad sami vyzkoušet...



Další typy ikonových grafů



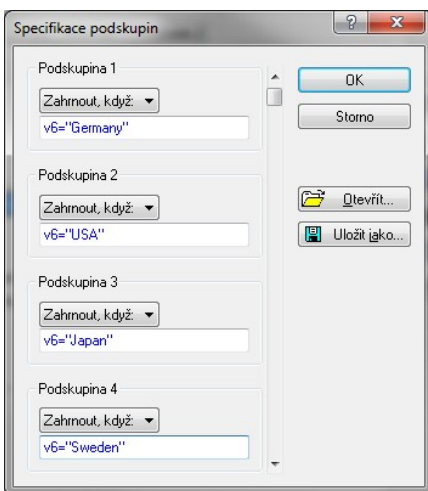
Poznámka: Pokud se zeptáte na rozdíl mezi Chernofovými tvářemi a ostatními ikonovými grafy, pak takový rozdíl tady opravdu je. Když nebudeme mluvit o faktu, že tváře jsou jistě vizuálně nejatraktivnější, máme zde další odlišnosti - tvářím se občas vytýká, že mohou být hodnoceny subjektivně podle toho, kdo zrovna obrázky posuzuje a které atributy obličeje považuje za důležité, které ne, a jak dokáže vyhodnotit odlišnosti. Na druhou stranu, tím, že jednotlivé části obličeje nemají řekněme stejnou důležitost (jako je tomu například u hvězdic), je možné vytvořit obličej tak, aby ty nejdůležitější znaky korespondovaly s tím, co považujeme ve tváři jako nejvýraznější (obecně se v této souvislosti zmiňuje jako velmi důležitá velikost očí a sklon obočí).

Příklad 2 (Auta)

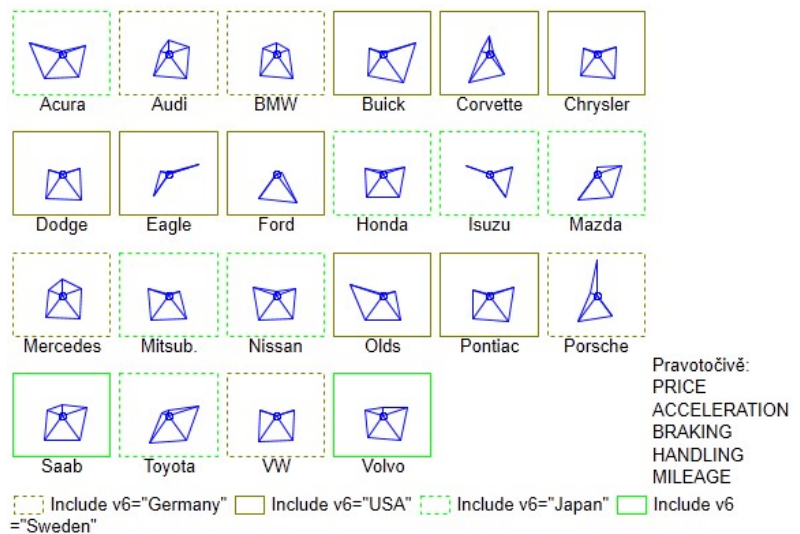
Ukažme si ještě jeden zajímavý příklad na datech o autech, kde se měřily jízdní i jiné vlastnosti aut (tato data jsou v příkladech programu *STATISTICA* pod názvem *Cars.sta*). Abychom si ukázali ještě další možnosti ikonových grafů, vytvořme si k pěti stávajícím proměnným s vlastnostmi novou proměnnou s příslušností auta k danému státu.

	1	2	3	4	5	6
	PRICE	ACCELERATION	BRAKING	HANDLING	MILEAGE	country
Buick	-0,614	1,689	0,933	-0,210	-0,154	USA
Corvette	1,235	-1,811	-0,494	0,973	-0,677	USA
Chrysler	-0,614	0,073	0,427	-0,210	-0,154	USA
Dodge	-0,706	-0,196	0,481	0,145	-0,154	USA
Eagle	-0,614	1,218	-4,199	-0,210	-0,677	USA
Ford	-0,706	-1,542	0,987	0,145	-1,724	USA
Honda	-0,429	0,410	-0,007	0,027	0,369	Japan
Isuzu	-0,798	0,410	-0,061	-4,230	1,067	Japan
Mazda	0,126	0,679	-0,133	0,500	-1,724	Japan
Mercedes	1,051	0,006	0,120	-0,091	-0,154	Germany

Vytvoříme ikonový graf, ale navíc v záložce details nastavíme podmínky pod tlačítkem *Označit ikony* následujícím způsobem:



Ikony, které přísluší daným kategoriím, budou nyní navíc orámovány podle toho, jestli splňují danou podmínku nebo nikoli, což může posloužit pro lepší orientaci.



Vidíme, že nejvíce odlišnými od ostatních aut jsou Porsche, Eagle a Isuzu, ty se vymykají, při pohledu do dat zjistíme, že například u Porsche je to kvůli ceně a zrychlení. K sobě velmi podobná jsou auta Mercedes, BMW a Audi nebo například Honda a Volkswagen.

Závěrem

Tímto příkladem bychom ukončili exkurz do ikonových grafů, doufáme, že pro Vás bude tato vizualizační technika přínosná a současně přejeme hodně krásných tváří, nejen Chernoffových.